SRI International

Semiannual Report • March 1995

AUDITION AND VISION IN VIRTUAL ENVIRONMENTS

Prepared by:

Thomas P. Piantanida, Principal Investigator
Medical Technology Laboratory

Prepared for:

Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Attn:Code ONR 342PS

SRI Project 5729

Contract N00014-94-C-0097

Original contains color plates: All DTIC reproductions will be in black and white

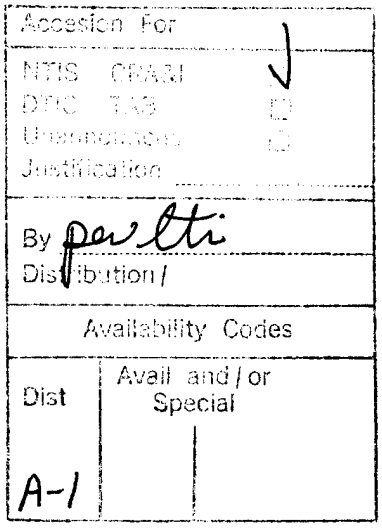
Approved by:

Ajit Shah, Director
Medical Technology Laboratory

Donald L. Nielson, Vice President
Computing and Engineering Sciences Division

333 Ravenswood Avenue • Menlo Park, CA 94025-3493 • (415) 326-6200 • FAX: (415) 326-5512 • Telex: 334486

CONTENTS



ILLUSTRATIONS

SYNOPSIS

During the first year of this project, we configured the virtual-environment (VE) system to perform target-detection tests using time-to-detection and pathlength-to-detection metrics. We also configured the SRI dual-Purkinje-image Eyetracker to examine eye movements induced by watching moving images at various image-update rates, as occurs in VEs.

Our target-detection studies disclosed performance differences between real and VEs that suggest that traditional relative and absolute visual orientation cues may not provide beneficial information in VEs. For some subjects, inclusion of orientation cues (achromatic and chromatic longitude and latitude grids) impaired target-detection performance, while in others it improved performance, but not significantly. We suspect that the orientation cues were not only too simplistic to allow most subjects to construct a veridical cognitive map of the VE, but also they required some subjects to restrict their range and rate of head movements to avoid simulator sickness. We plan to test this assumption by providing more realistic orientation cues, such as architectural features (e.g., doors and windows) and cultural features (e.g., furniture and prints), in our VEs.

Not surprisingly, target-detection performance, with and without orientation cues, showed a clear first-order effect of field-of-view size. Performance varied inversely with field size. Our studies show that performance is quite poor at small fields of view, improves significantly at field sizes exceeding about 30 degrees, and is still not asymptotic at 100-degrees field of view. These results suggest that the current method of improving resolution in head-mounted displays (HMD)—by reducing field of view—severely affects the user's ability to find objects in the VE.

Initial observations with acoustic icons as absolute orientation cues in sparse VEs suggest that the auditory system is not particularly adept at generating cognitive maps of virtual spaces. We first used a regular tetrahedron of acoustic icons surrounding the subject's head to provide the subject with direction-of-gaze information, on the assumption that target-detection performance would be enhanced by the addition of acoustic orientation cues. In pilot studies, subjects reported that the orientation information provided by the acoustic icons could not be assimilated without conscious effort, which impaired target-detection performance. In addition, subjects experienced large discrepancies between their presumed orientation in the laboratory and their actual orientation. We plan to study the effect of acoustic icons as target-bound orientation cues.

We also plan to expand our target-detection studies into the realm of discrimination by initiating a series of studies to examine target-discrimination performance in a low-resolution VE, such as can be generated with current consumer-quality HMDs. These studies will be an adjunct to our variable-resolution studies that we plan to conduct on our high-resolution stereoscope. The stereoscope will use two high-resolution monitors to present stereoscopic images to the subject. Image resolution can be reduced systematically by selecting subsets of display pixels to generate stimulus pixels. For example, a stimulus pixel might consist of 4 or 9

or 16 display pixels that can only vary in brightness as a unit. The stimulus, then, will consist of essentially a spatial-frequency-filtered version of the high-resolution display image.

On another front, we expect to begin running naive subjects in our eye-tracker study within the next 2 weeks. Our pilot studies, conducted with sophisticated subjects, have provided interesting preliminary results. These pilot studies examine the eye movements of persons watching moving stimuli on a visual display that can be updated at a variable rate from 1 to 60 Hz. Consistent with our hypothesis of saccadic eye movements at low update rates and pursuit movements at high update rates, the preliminary data show a band of update rates where neither saccades nor smooth pursuit accurately tracks the moving stimulus. We have updated our analog-to-digital input/output system so that it can capture and analyze eye rotation signals and their first derivatives from the eyetracker and can acquire stimulus position information from the visual display. This will allow us to calculate the magnitude of retinal-image error, which we believe will correlate with subjective reports of visual discomfort and simulator sickness.

VISUAL ORIENTATION CUES

In this series of experiments, subjects were required to locate a specific target within a VE that may or may not contain relative and absolute visual orientation cues, and which they viewed through virtual apertures of variable size. The VE consisted of a uniform, white virtual sphere with a diameter of 40 feet, upon which could be superimposed a spherical grid of latitude and longitude lines; these were all black in the case of relative orientation cues, and included colored longitude lines at the major compass points in the case of absolute orientation cues. Longitude lines were provided at 30-degree intervals, and latitude lines were presented at the equator and at ±15-degree increments to ±45 degrees from the equator. The target and any distractors were constrained to appear within latitudes ±45 degrees, at any azimuth.

The target was always a red square subtending approximately 200 minutes of arc on a side. Distractors consisted of four red discs and four green squares that all subtended the same visual angle as the target. The virtual apertures, which subtended 14, 28, 41, or 53 degrees of visual angle, had surrounds that matched the virtual sphere perfectly. Thus, unless virtual features, such as latitude and longitude lines, or the target or distractors, were present within the aperture, the subject could not tell that he was viewing the virtual sphere through an aperture. The fifth field-of-view condition was a no-aperture condition, in which the subject viewed the entire 60-degree-vertical by 100-degree-horizontal image of the virtual sphere. Figure 1 illustrates the subject's view of the virtual sphere.

The virtual sphere and all virtual features were produced on the binocular liquid-crystal displays (LCDs) of the VPL EyePhones. Pixel pitch of the LEEP-Optics-magnified images of the LCDs was approximately 13 minutes of arc. Thus, the upper limit of spatial frequencies that could be presented in the EyePhones was approximately two cycles per degree.

Head tracking was accomplished with a Polhemus magnetic tracker that sampled head position and orientation at 60 Hz. Tracker range was only accurate over a 3-foot radius, so the

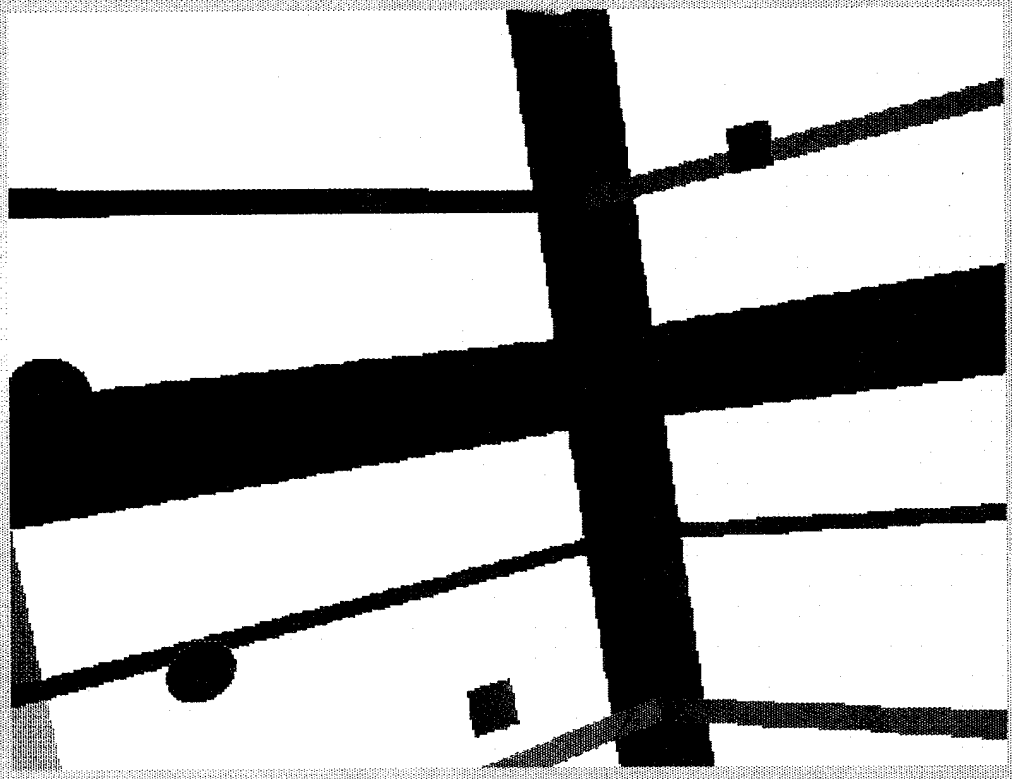


Figure 1a
View of target and three distractors with grid and no aperture

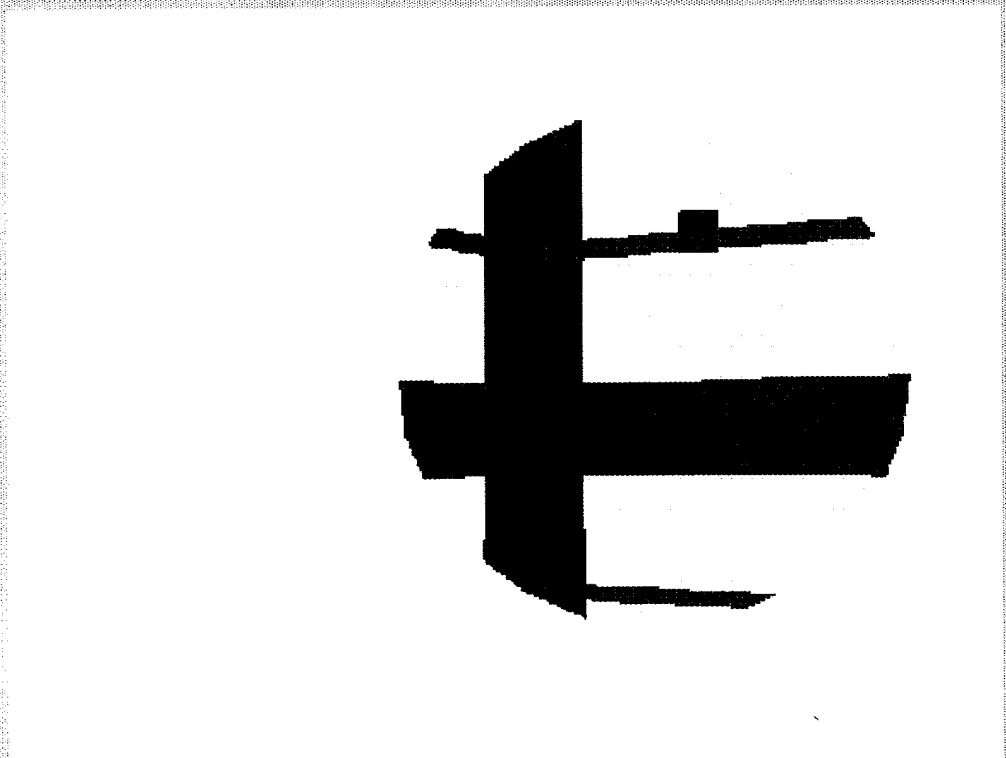


Figure 1b
View of target with grid and 28 degree aperture

subject was constrained to this volume. To aid the subject, a 5-inch plastic disc was affixed to the floor directly under the Polhemus emitter. The subject could maintain his position relative to the emitter by feeling the disc with his foot. Pitch and yaw were measured at 60 Hz and recorded.

The subject wore a Virtex Cyber Glove to communicate with the computer that ran the experiment. By clenching his fist, the subject could signal that he had found the target. The computer verified that the target was visible within the aperture and recorded the response as a hit; otherwise the response was rejected.

Under each of the 15 conditions of the experiment (five aperture conditions and three orientation conditions), time between successive hits and pathlength between successive hits were recorded for 32 consecutive presentations. The last 30 responses were recorded and analyzed. Detection time was recorded in head-tracker polling units—i.e. 1/60 second. Pathlength was calculated as the sum of the number of degrees of pitch and the number of degrees of yaw between successive hits. Pathlength-to-detection was selected as an alternative metric in these experiments because pilot studies indicated that the presence of orientation cues in the virtual sphere caused some subjects to slow their head movements, apparently to reduce vertigo. Figure 2 shows the group mean time-to-detection and pathlength-to-detection data from all subjects at each of the aperture sizes for each of the orientation-cue conditions.

As we reported previously, the data are approximately linear in $\log_{10}$ time/pathlength-$\log_{10}$ aperture diameter coordinates. Replotted data for all subjects are shown in Figure 3. The slopes of the functions show two clusters: one for time-to-detection and another for pathlength-to-detection, which differ in slope. Although the pathlength and time data are only arbitrarily related, the fact that mean pathlength-to-detection and mean time-to-detection functions differ with respect to aperture diameter suggests that they measure different aspects of search strategies adopted by the subjects. We plan to examine this difference more closely in future studies. The correlation coefficients for the logarithmic curve fit vary from a low of 0.722 to a high of 0.888, which suggests that much of the variance in the data set is accounted for by the fit of the log function.

The reduction in the rate of head-movement velocities under the relative- and absolute-orientation conditions confounded the time-to-detection metric for some subjects. Mean time-to-detection curves are shown in Figure 4 for one subject who showed a reduction in head-movement velocity. Figure 5 shows similar data for a second subject. Other subjects, however, showed no apparent reduction in head-movement velocity under the relative- and absolute-orientation conditions. Mean time-to-detection curves are shown for one of these subjects in Figure 6 and for another in Figure 7. Note that in Figures 4 and 5, many or all of the normalized time-to-detection points fall below zero, indicating that the relative and absolute orientation cues impact target-detection performance negatively. In Figures 6 and 7, on the other hand, the normalized time-to-detection functions tend to fall above the zero effect line, indicating an enhancement of target-detection performance in the presence of relative- and absolute-orientation cues.

The data in Figures 8 through 11 show that the relative- and absolute-orientation cues tended to exert their major effects, either enhancement or decrement, at the smaller aperture sizes. The data clearly converge at larger aperture sizes.

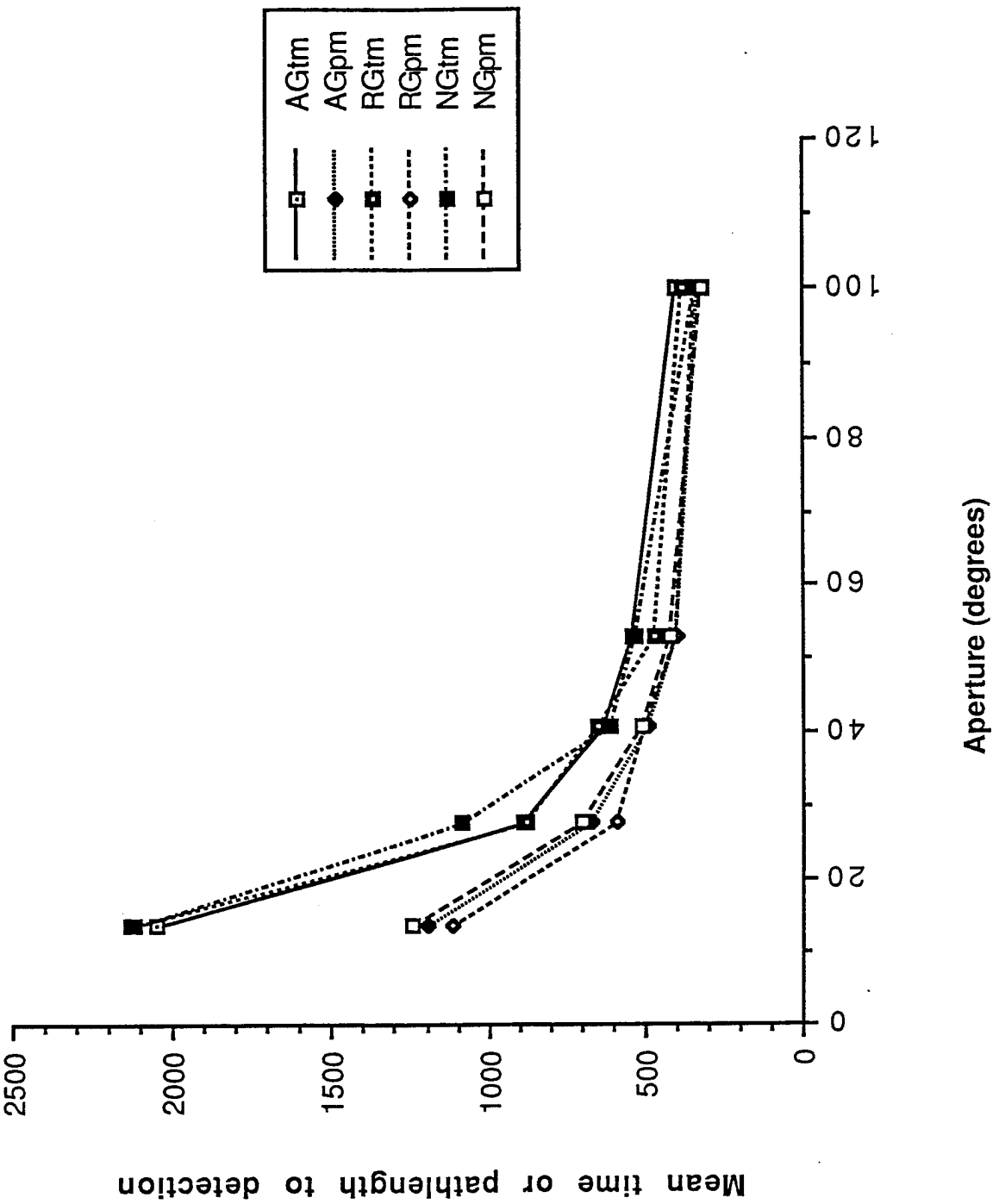


Figure 2. Grand mean time-to-detection (in sec/60 time units) and pathlength-to-detection (in cumulative degrees of pitch and yaw) as a function of aperture diameter and orientation-grid type. AGtm is the grand mean time-to-detection with an absolute-orientation grid. AGpm is the grand mean pathlength-to-detection with an absolute-orientation grid. RGtm is the grand mean time-to-detection with a relative-orientation grid. RGpm is the grand mean pathlength-to-detection with a relative-orientation grid. NGtm is the grand mean time-to-detection without an orientation grid. NGpm is the grand mean pathlength-to-detection without an orientation grid.

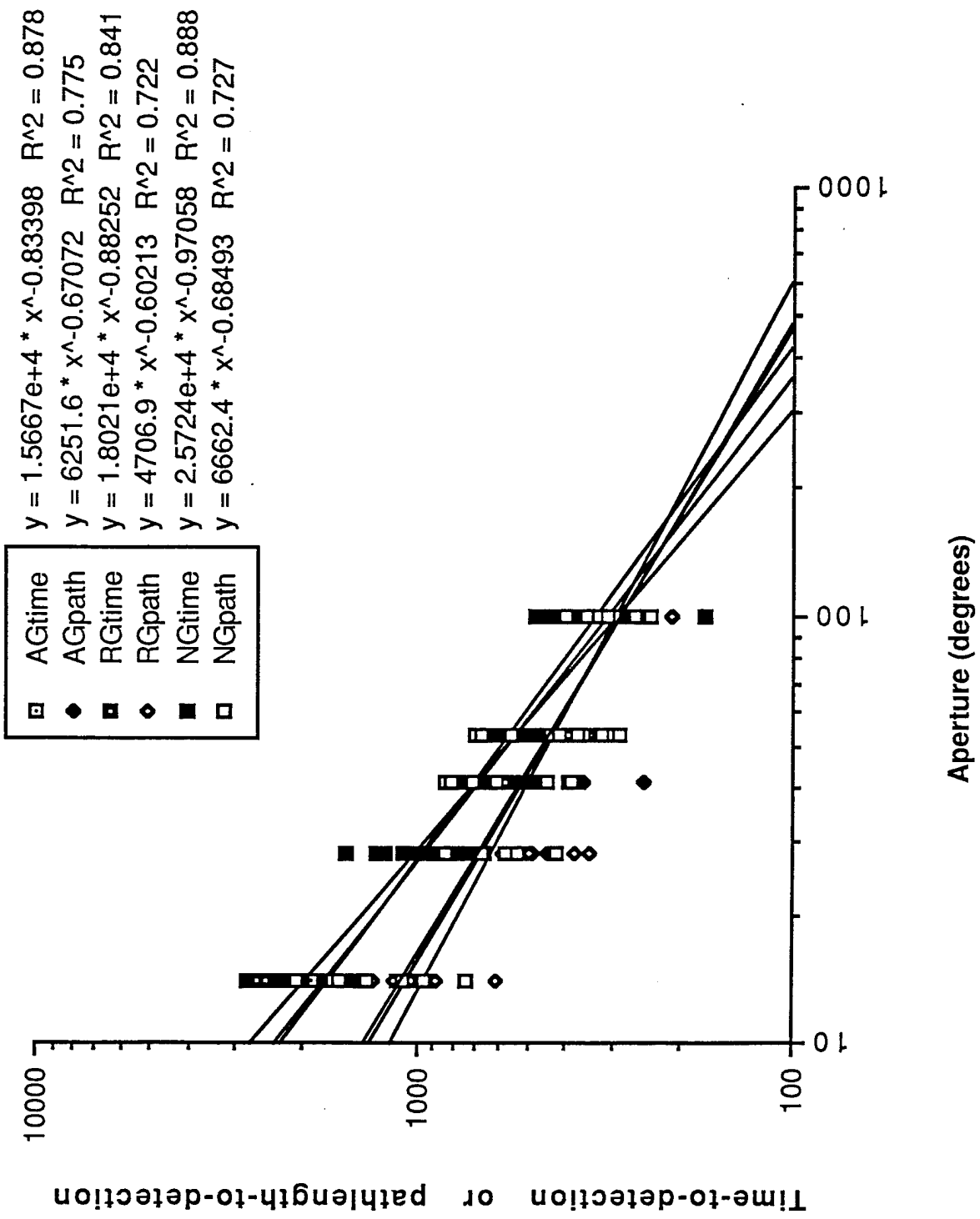


Figure 3. Individual subject mean times-to-detection and pathlengths-to-detection as a function of aperture diameter and orientation-grid type plotted in log-log coordinates. Logarithmic regression lines and correlation coefficients are shown for conditions AGtime (time-to-detection with an absolute-orientation grid), AGpath (pathlength-to-detection with an absolute-orientation grid), RGtime (time-to-detection with a relative-orientation grid), RGpath (pathlength-to-detection with a relative-orientation grid), NGtime (time-to-detection without an orientation grid), and NGpath (pathlength-to-detection without an orientation grid).

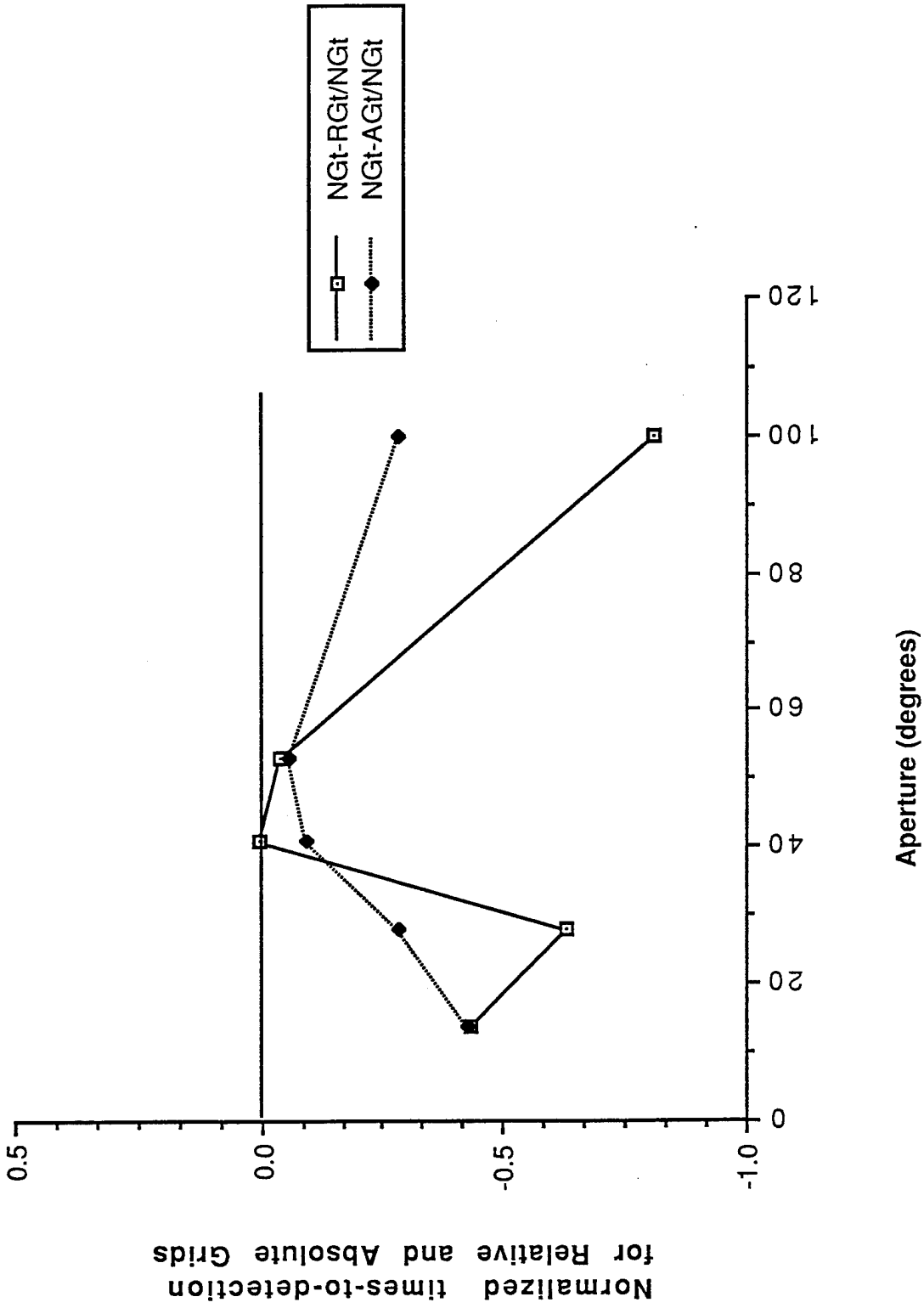


Figure 4. Normalized mean time-to-detection data for relative- and absolute-orientation-grid conditions as a function of aperture diameter for subject RL. NGt-RGt/NGt is the normalized difference between relative-orientation-grid and no-grid means. NGt-AGt/NGt is the normalized difference between absolute-orientation-grid and no-grid means. All of the values are below zero, indicating a performance deficit for both relative- and absolute-orientation-grid conditions.

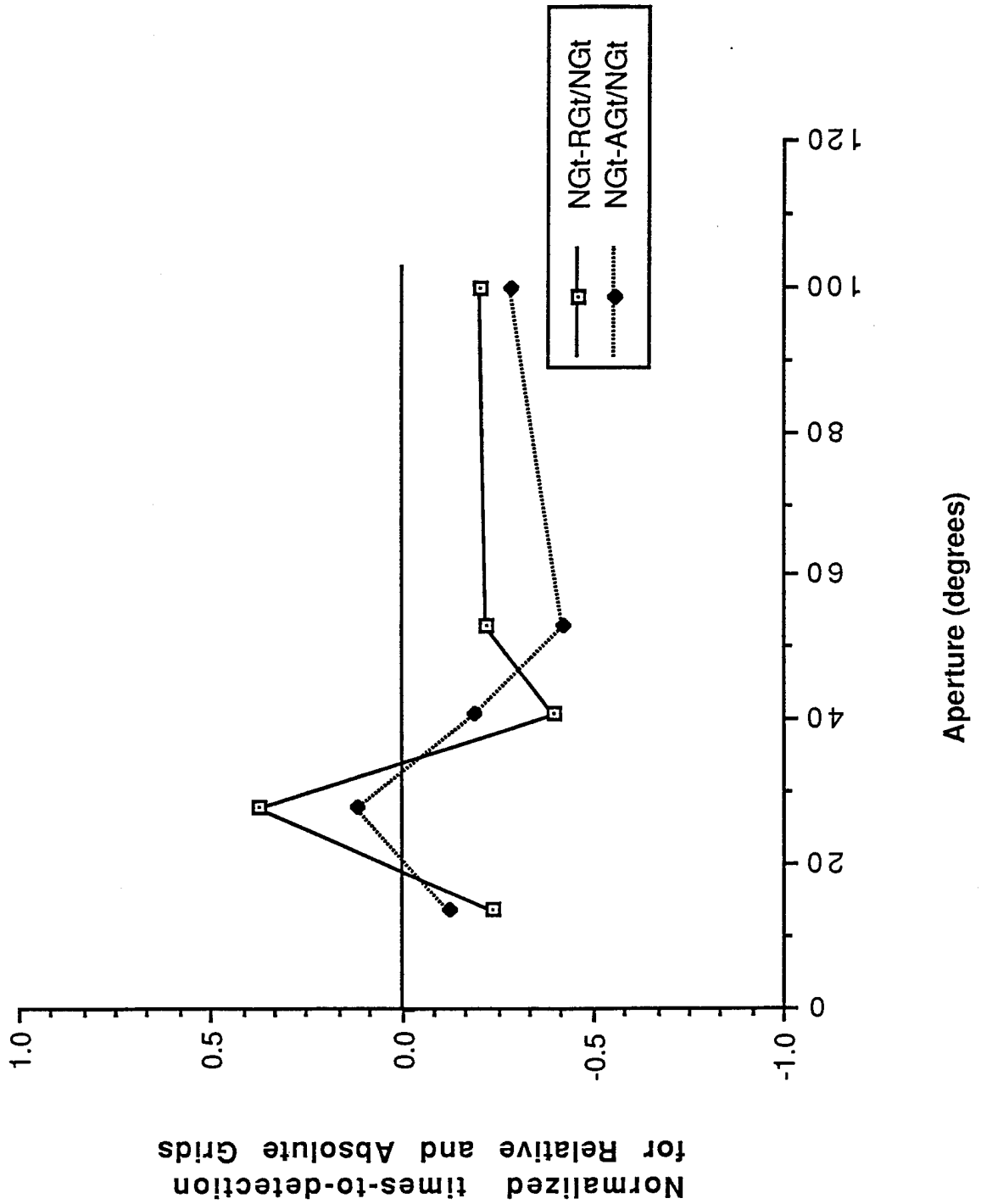


Figure 5. Normalized mean time-to-detection data for relative- and absolute-orientation-grid conditions as a function of aperture diameter for subject PS. Most means fall below zero, indicating a performance deficit for both relative- and absolute-orientation-grid conditions. Legend as in Figure 4.

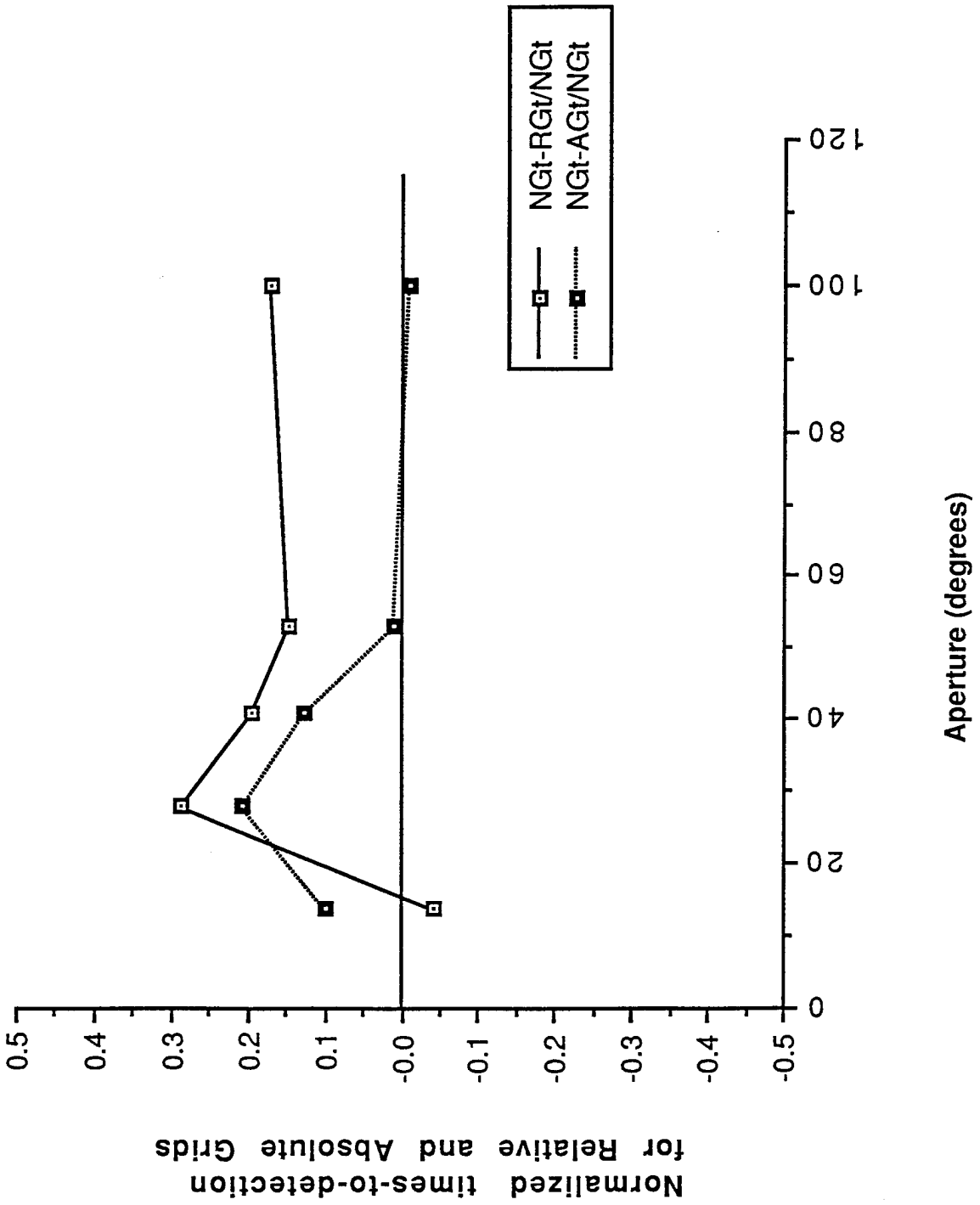


Figure 6. Normalized mean time-to-detection data for relative- and absolute-orientation-grid conditions as a function of aperture diameter for subject DD. Most means fall above zero, indicating enhanced performance for both relative- and absolute-orientation-grid conditions. Legend as in Figure 4.

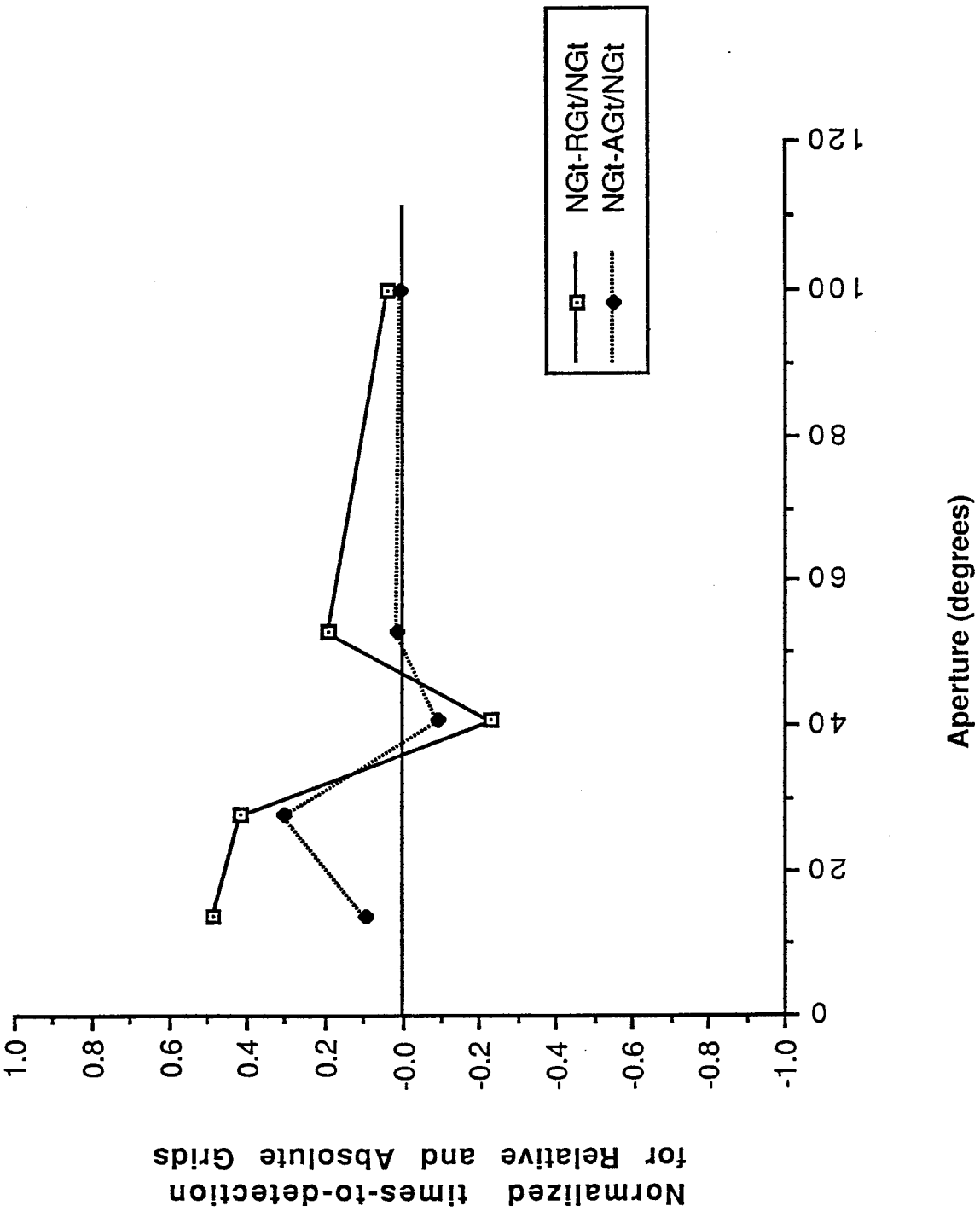


Figure 7. Normalized mean time-to-detection data for relative- and absolute-orientation-grid conditions as a function of aperture diameter for subject JK. Most means fall above zero, indicating enhanced performance for both relative- and absolute-orientation-grid conditions. Legend as in Figure 4.

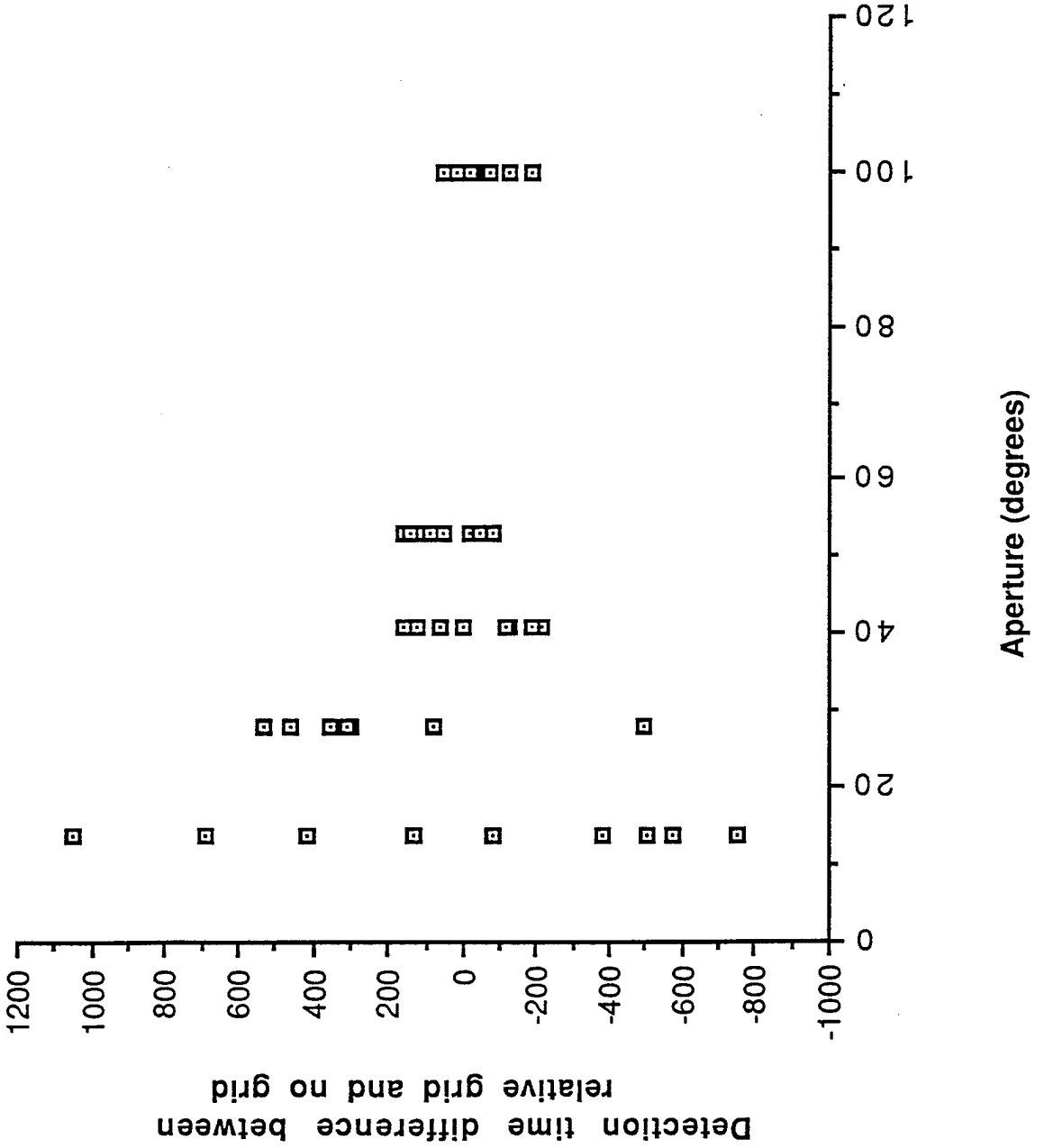


Figure 8. Differences in mean time-to-detection between relative-orientation-grid and no-orientation-grid conditions plotted for all subjects as a function of aperture diameter. Positive values indicate performance enhancement and negative values indicate performance deficits.

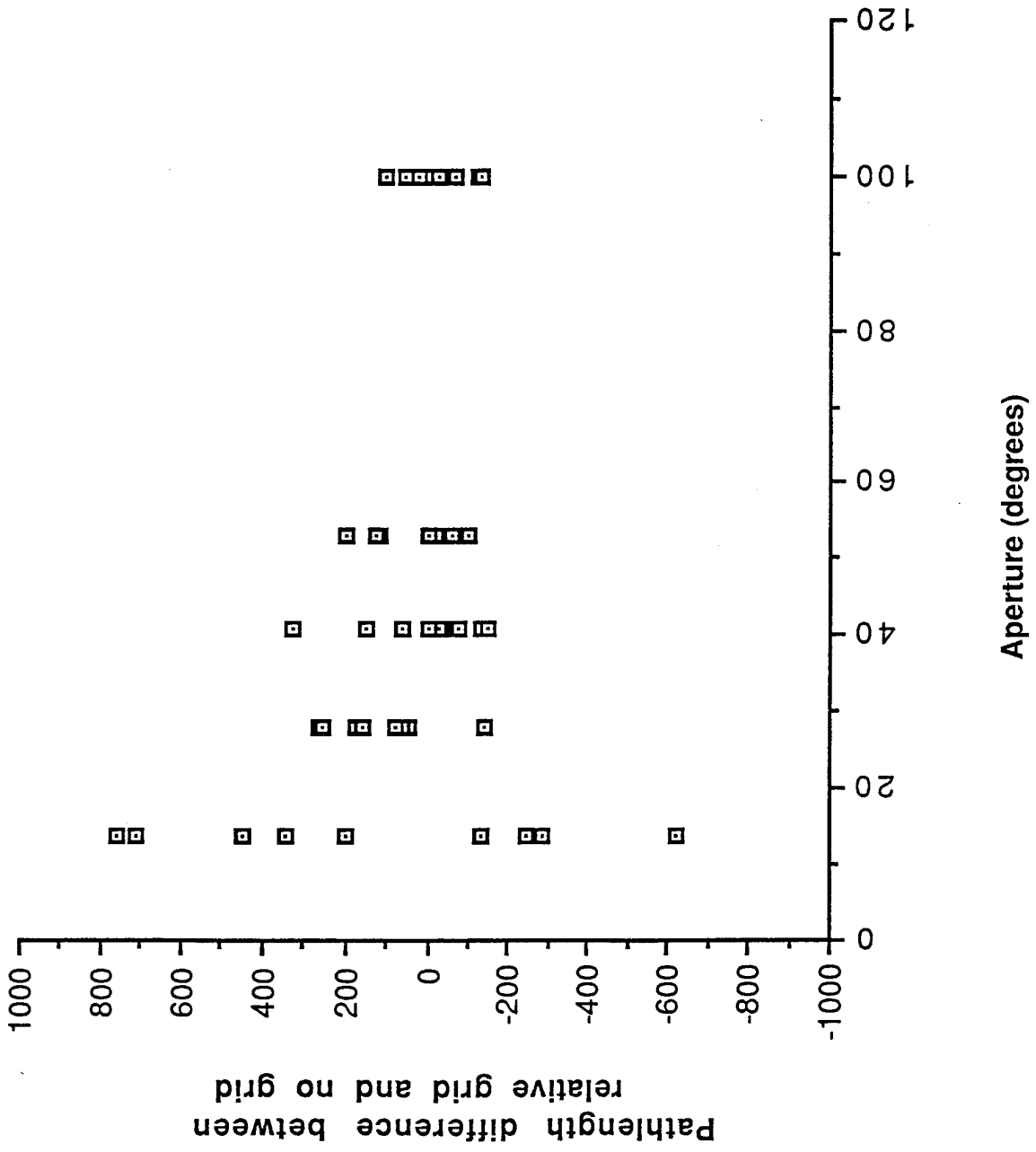


Figure 9. Differences in mean pathlength-to-detection between relative-orientation-grid and no-orientation-grid conditions plotted for all subjects as a function of aperture diameter. Positive values indicate performance enhancement and negative values indicate performance deficits.

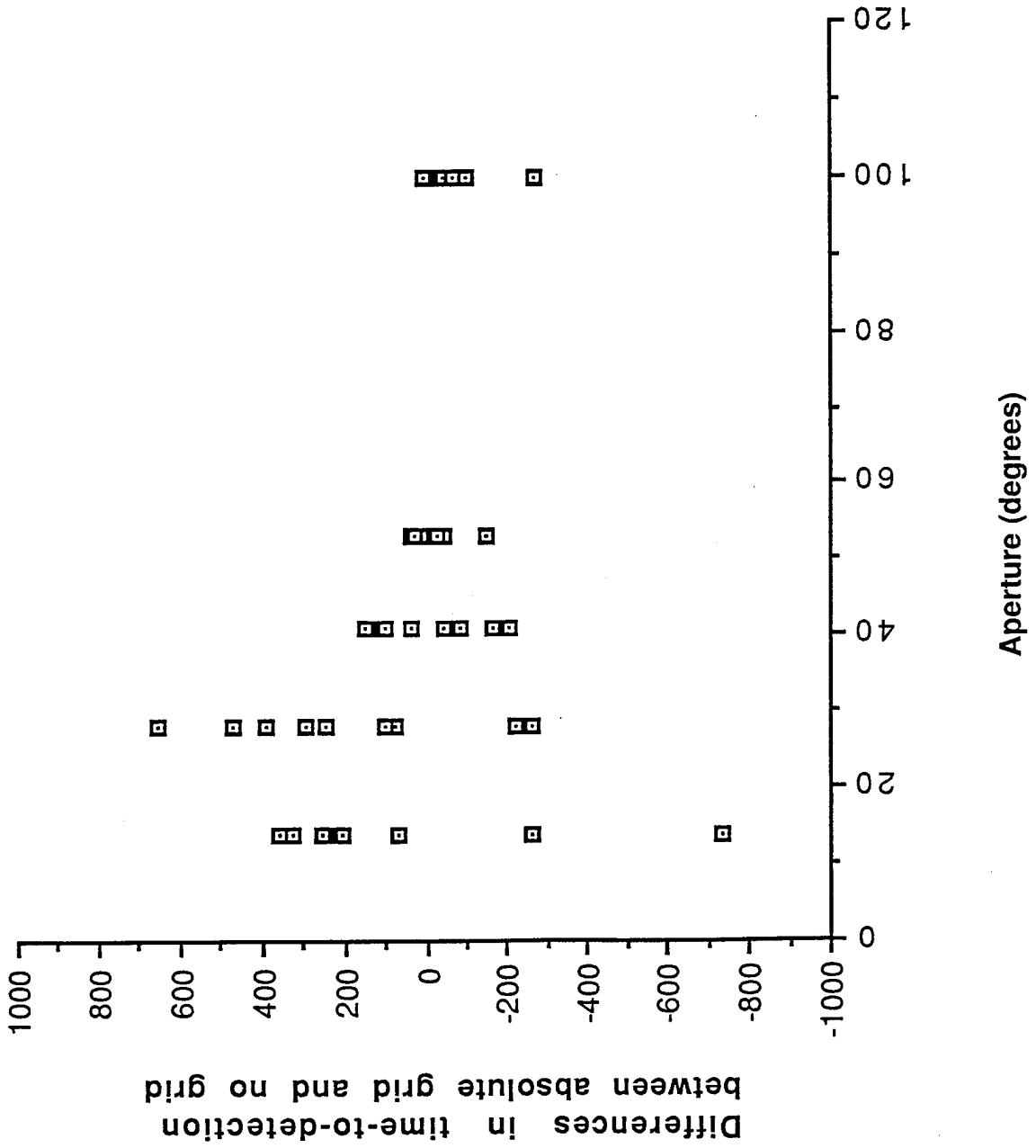


Figure 10. Differences in mean time-to-detection between absolute-orientation-grid and no-orientation-grid conditions plotted for all subjects as a function of aperture diameter. Positive values indicate performance enhancement and negative values indicate performance deficits.

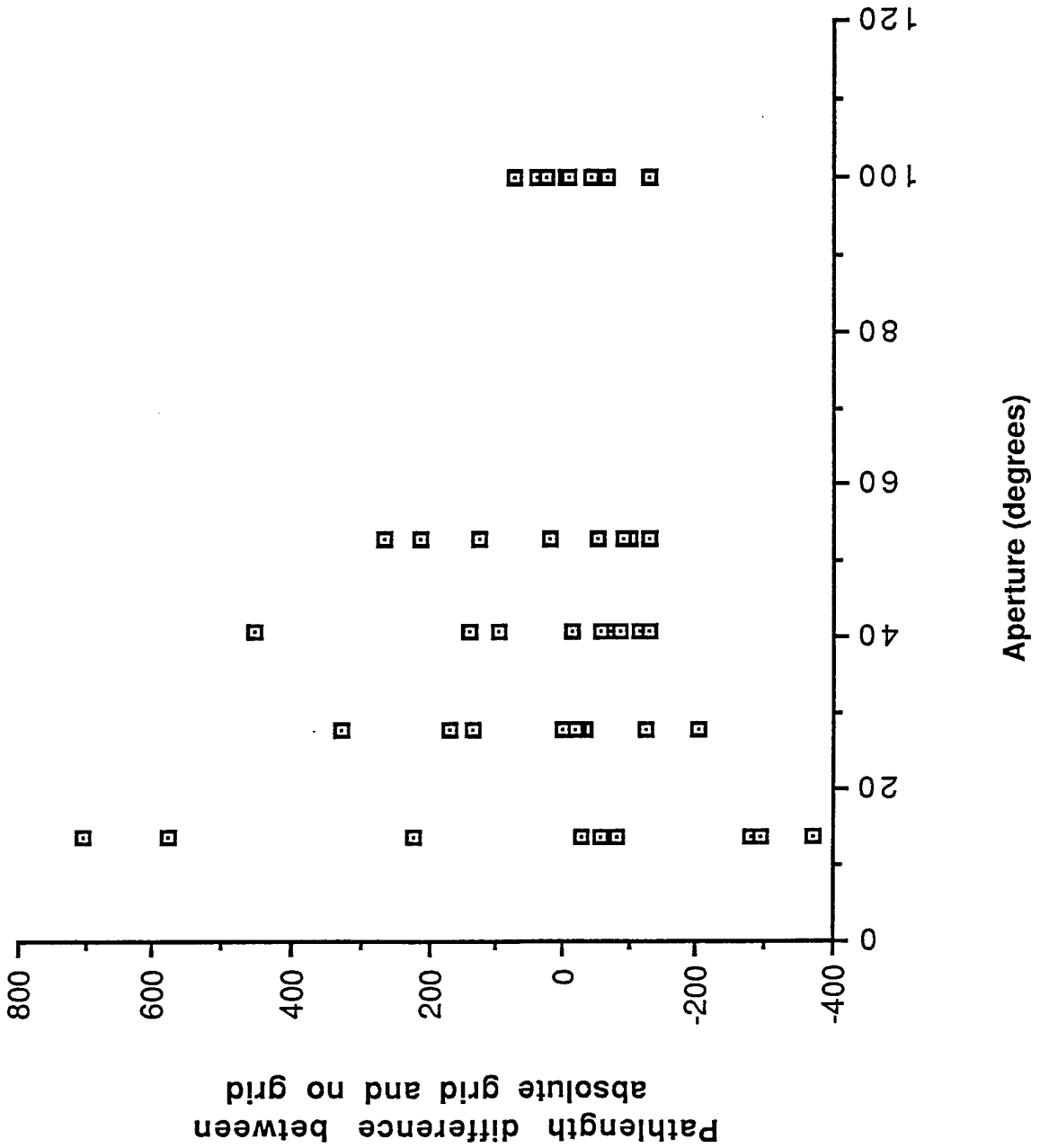


Figure 11. Differences in mean pathlength-to-detection between absolute-orientation-grid and no-orientation-grid conditions plotted for all subjects as a function of aperture diameter. Positive values indicate performance enhancement and negative values indicate performance deficits.

Data from all subjects were analyzed using several order statistics. The Friedman Test for Matched Groups was used to evaluate an effect attributable to absolute or relative orientation cues.

Using the Friedman Test, $X2 = \frac{12}{N.C(C+1)} [* T2]- N.C(C+1)$

where N is the number of subjects, C is the number of conditions, and T is the sum of ranked mean time-to-detection values summed over all apertures for each subject. Results are shown in Table 1.

Table 1

RESULTS OF FRIEDMAN TEST

	Condition					
	Absolute Grid		Relative Grid		No Grid	
Subject	Value	Rank	Value	Rank	Value	Rank
DD	4444	2	4432	3	4994	1
JK	4416	2	3395	3	4972	1
LC	5243	3	5477	2	5499	1
MA	4589	2	4030	3	4947	1
MP	3678	3	4579	1	4215	2
PS	4588	2	4621	1	4088	3
RL	4834	2	5195	1	3738	3
RR	4268	3	4990	2	5059	1
SG	4052	2	3689	3	4628	1
Σ		21		19		14

$\chi^2 = 2.878$, df = 2, $p \leq 0.250$

The results failed to reach statistical significance, indicating that the relative and absolute orientation cues did not significantly affect time-to-detection performance on a target-detection task.

The Wilcoxon Match Pairs Signed Rank Test was used to assess whether orientation cues per se had any significant effect on target-detection performance. A one-tailed test indicated that $p > 0.05$, again suggesting that, while provocative, the data do not support the hypothesis that visual orientation cues affect target-detection performance in VEs.

ACOUSTIC ORIENTATION CUES

Pilot studies were conducted to assess the effectiveness of localized acoustic sources as orientation cues in a VE. The studies were conducted using the same apparatus as in the case of the visual orientation cues, except that no visual orientation cues were ever present, and a regular tetrahedron of virtual acoustic icons was positioned to surround the subject's head. This tetrahedron consisted of one acoustic source positioned directly over the subject's head and a triangular array of three acoustic sources positioned at about waist level to the subject. All acoustic sources were generated by the Crystal River Convolvotron, which creates spatialized sounds by recreating the transit delays and head-related transfer functions that the auditory system uses to localize sound sources.

Initially, acoustic icons had clearly identifiable complex signatures and were presented continuously. This proved to be quite confusing to the subject. By changing the icons to intermittent acoustic trains, the sounds were reportedly localized more readily. However, the perceived stability of the icons—that is, the perception that each was emanating from a fixed point in space—was never robust enough to allow them to be used as fixed reference points. We plan to continue exploration of acoustic icons as anchoring stimuli. Perhaps the subject could orient more accurately if the acoustic icons were placed on the equator at the cardinal compass points, or if there were fewer icons.

UPDATE RATE STUDIES

We have assembled an apparatus that will allow us to examine the eye movements of subjects who are viewing moving stimuli that have update rates controlled by the experimenter. The apparatus consists of a monitor on which we present a vertical bar stimulus that oscillates across the screen at a selected speed and selected update rate, an SRI dual Purkinje-image Eyetracker that samples the subject's eye position and rotation at several kilohertz with an accuracy of approximately one minute of arc, and a MacADIOS interface for digitizing and analyzing the eyetracker data. The monitor is viewed at a distance of approximately 6 feet, so that the screen subtends approximately 25 degrees of visual angle. Stimuli—controllable width and contrast vertical bars—oscillate across the face of the monitor at a fixed speed controlled by the experimenter. For example, a stimulus might consist of a 2-degree-wide vertical black bar with 80% contrast, which requires 8 seconds to complete one cycle of oscillation. With these stimulus parameters fixed, the experimenter can select a fixed update rate for presenting the bar stimulus. An update of 1 Hz, for example, would cause the bar to be presented only eight times during one complete oscillation, while the same stimulus updated at 30 Hz would be presented 240 times during one complete oscillation. The eyetracker recorded the subject's eye movements

as he viewed the moving stimulus. We expected that at higher update rates, such as 20 or 30 Hz, the subject would make smooth pursuit eye movements to follow the moving bar, while at very low update rates—1 to 2 Hz, for example—the subject would make saccadic eye movements exclusively. Our objective was to determine the band of frequencies within which neither smooth pursuit nor saccades allowed the subject to track the stimulus accurately.

Pilot data were recorded on a strip chart and evaluated visually. Eye movement recordings are shown in Figures 12 through 17 for update rates of 1, 4, 6, 15, 30, and 60 Hz. The second trace on the chart is the first derivative of the eye movement signals, which discriminates between saccades and smooth pursuit. These pilot data show that saccadic eye movements are made in response to stimuli updated at 1, 4, and 6, and perhaps even 15 Hz , and that smooth pursuit eye movements are made in response to stimuli updated at 60, 30, 15, and perhaps even 6 Hz. The frequency region of overlapping pursuit and saccadic eye movements is the most interesting, as this appears to be the range of update rates that contribute to disorientation and simulator sickness in VEs.

We have recently integrated our MacADIOS system for acquiring eye movement data (including blinks, which cause artifacts in the record) as well as position data from the monitor screen. These data will allow us to calculate retinal position error as a function of stimulus update rate. We believe that these data will provide insights into the perceptual effects of low update rate in VE systems.

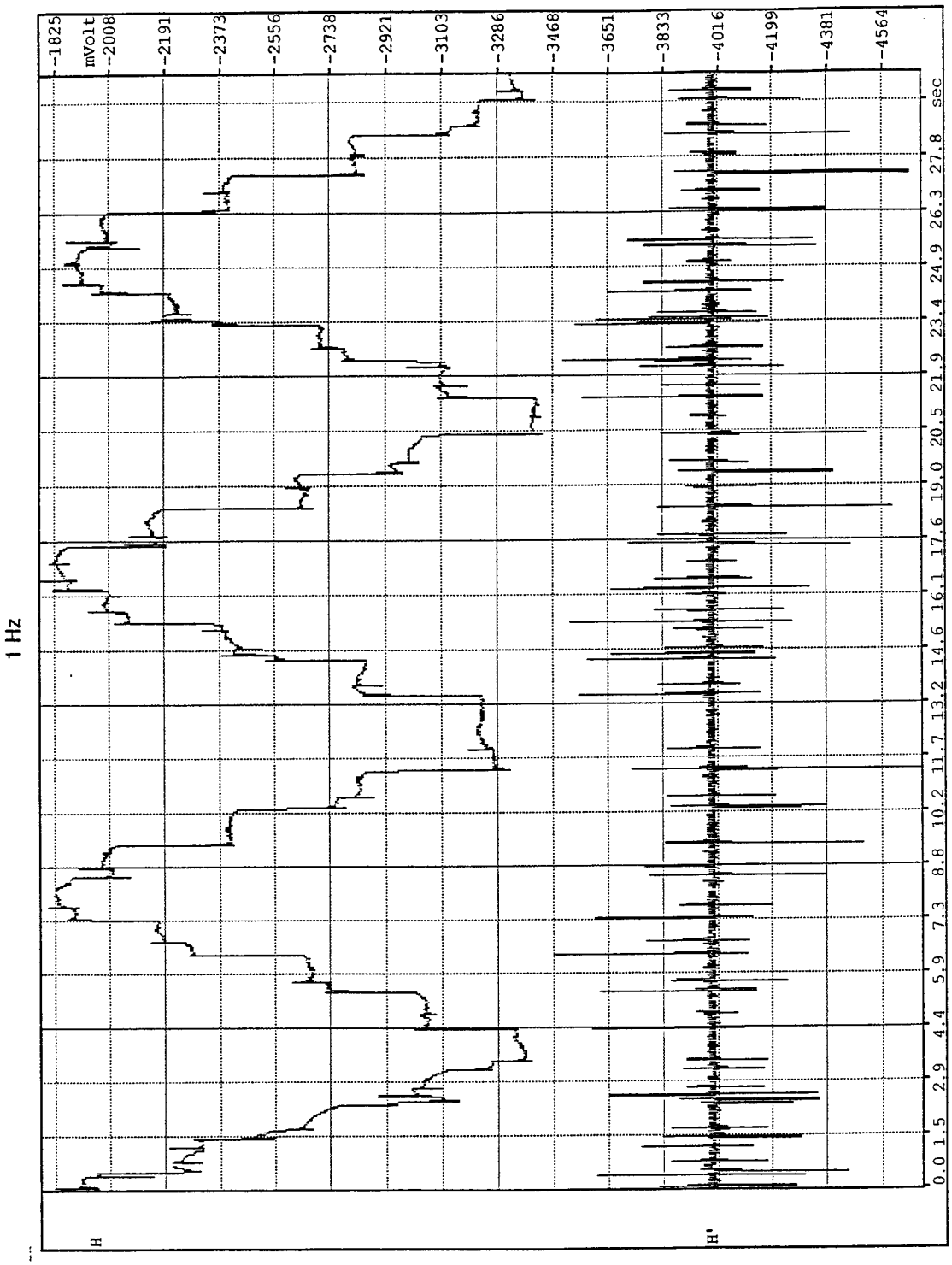


Figure 12. Data traces on a chart recorder showing eye position on the left and the first derivative of eye position on the right. These data were captured by the SRI dual-Purkinje-image Eyetracker as a subject viewed an image of a vertical black bar that oscillated horizontally with a period of 8 seconds and with a full excursion of 25 degrees against a white background. Plateaus in the eye-position trace indicate fixations and rapid lateral displacements of the trace indicate saccades. Spikes in the first-derivative trace show the frequency, direction, and magnitude of saccades. The position of the black bar was updated at 1 Hz.

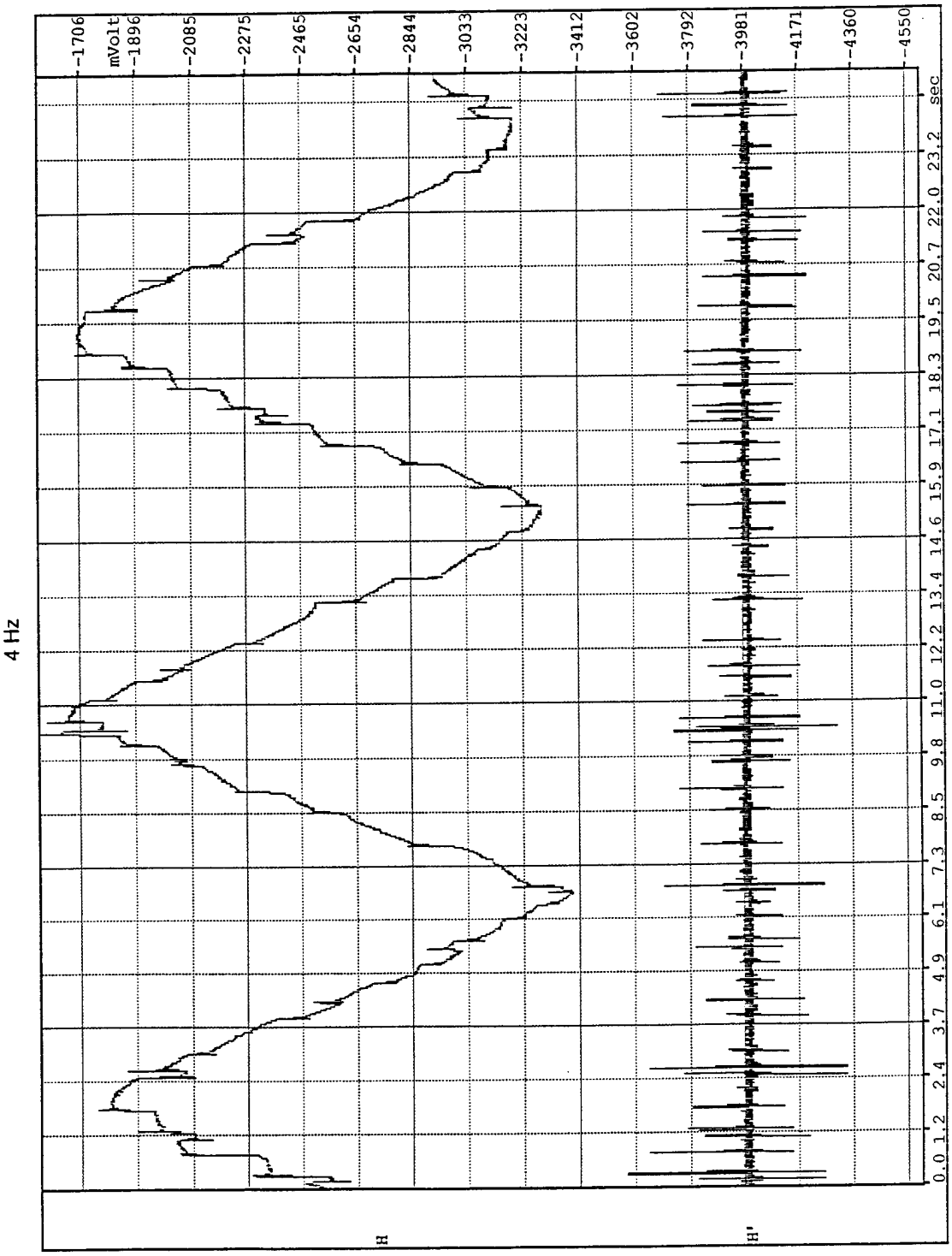


Figure 13. Same as in Figure 12, but with an update rate of 4 Hz. Saccades more frequent, but smaller than those seen in Figure 12. Curvature of the first-derivative baseline traces indicate that the bar is being tracked with smooth pursuit movements of the eyes.

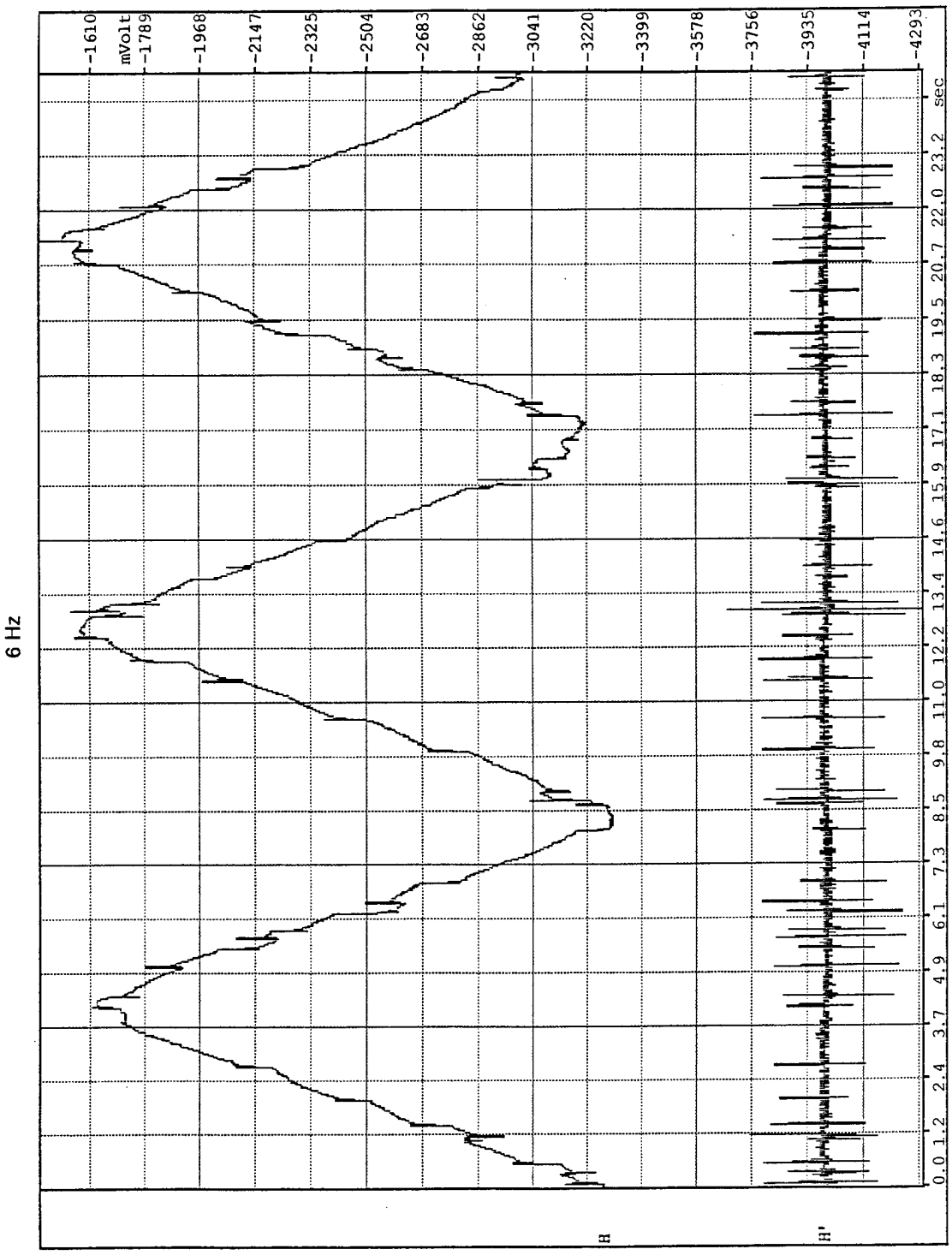


Figure 14. Same as in Figure 12, but with an update rate of 6 Hz. Saccades more frequent, but smaller than those seen in Figures 12 and 13. Increased curvature of the first-derivative baseline trace indicates an increase in smooth-pursuit eye movements, although saccades are still present.

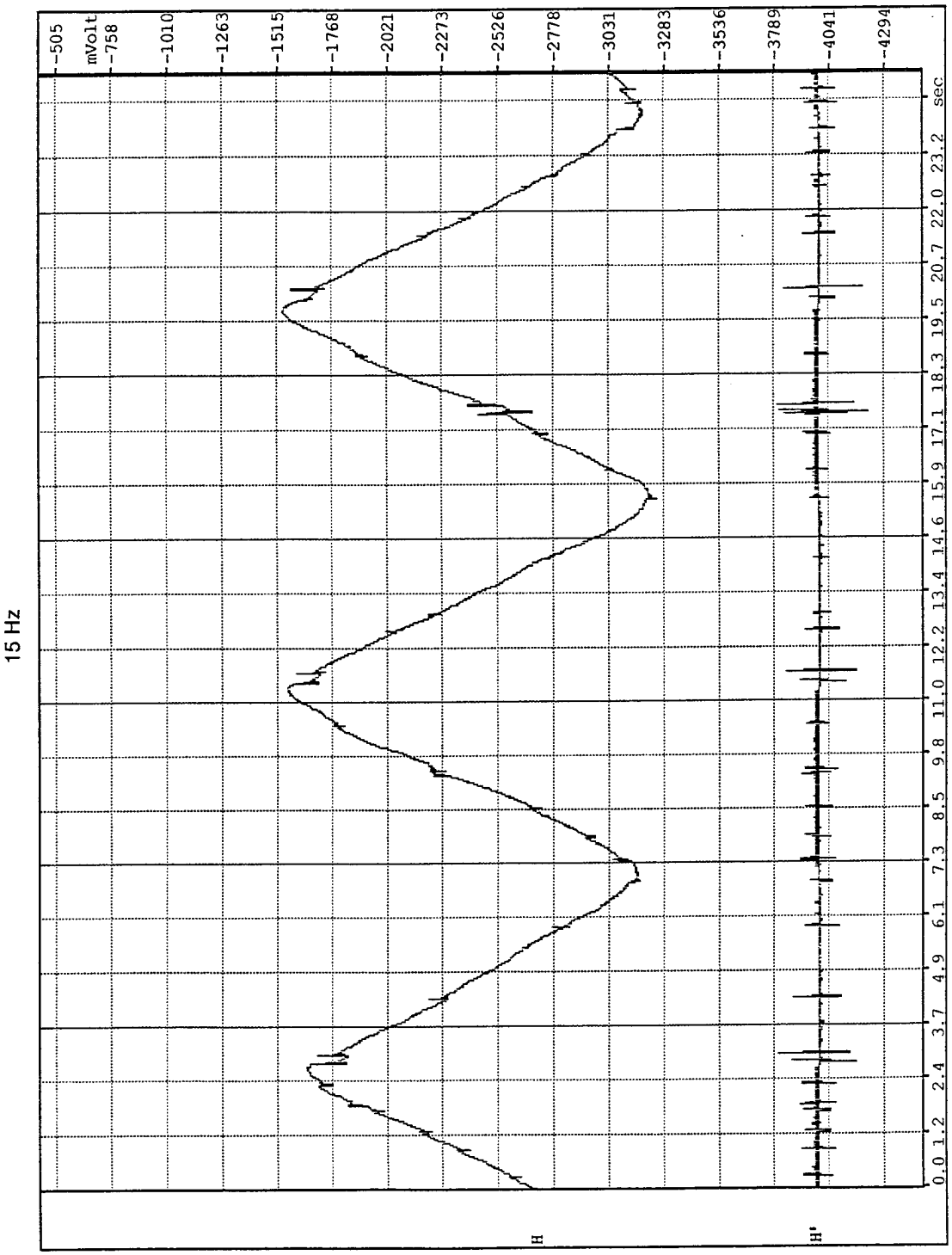


Figure 15. Same as in Figure 12, but with an update rate of 15 Hz. Saccades are still present, but of small amplitude and sporadic.

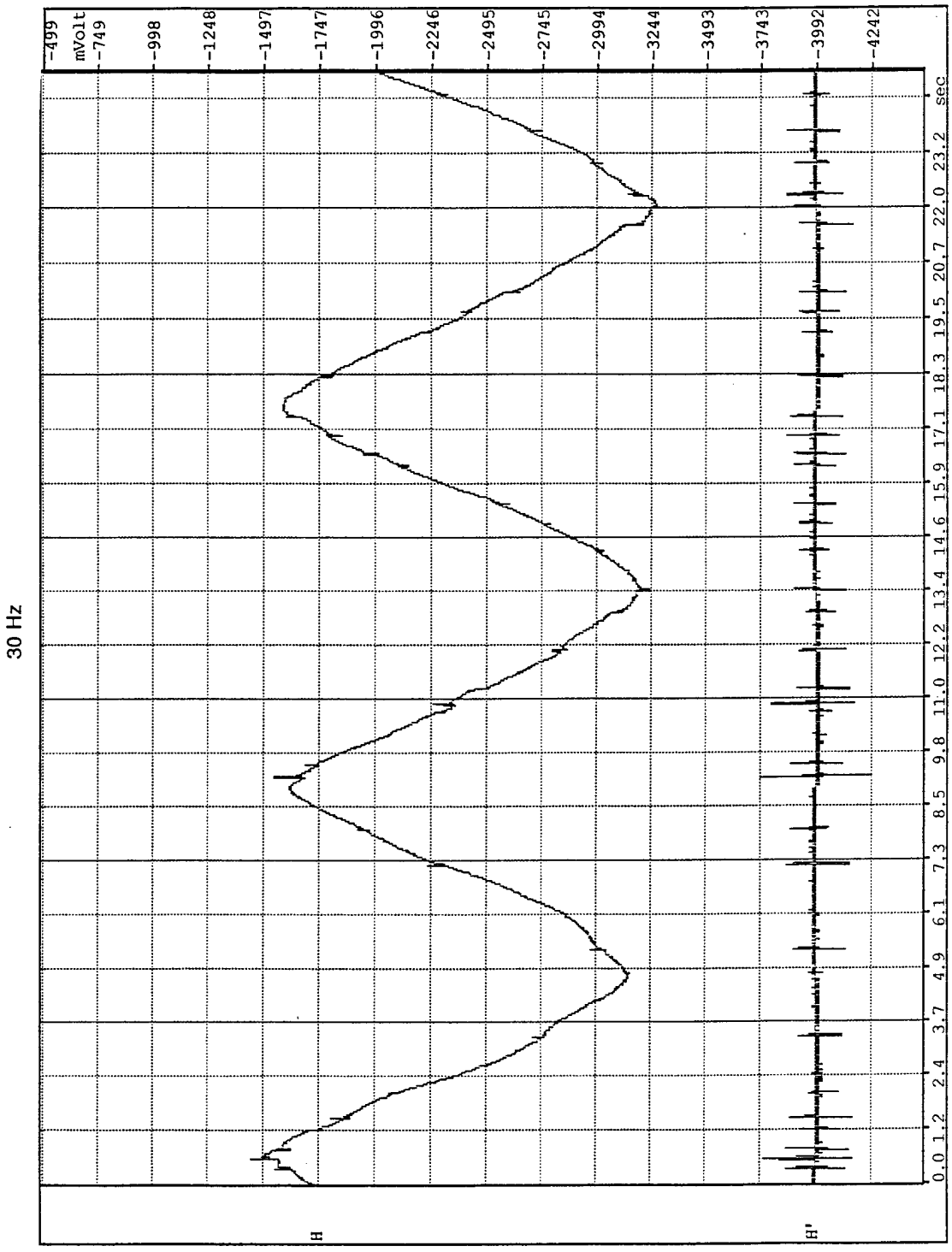


Figure 16. Same as in Figure 12, but with an update rate of 30 Hz. The traces show almost exclusively smooth-pursuit eye movements with little evidence of saccades except at the extremes of the bar oscillations, and an occasional corrective saccade.

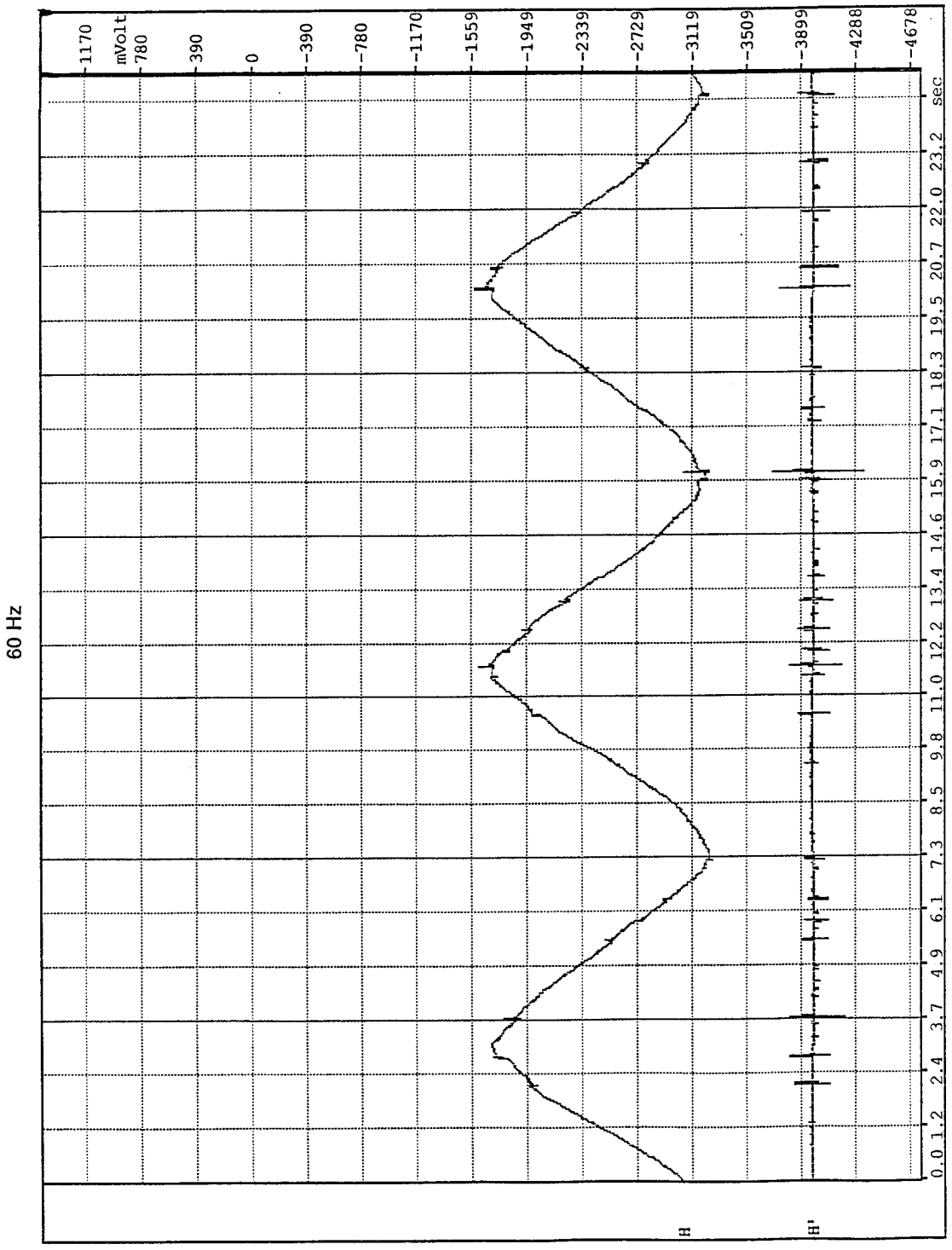


Figure 17. Same as in Figure 12, but with an update rate of 60 Hz. The traces show smooth-pursuit eye movements with an occasional corrective saccade.